

цифровой  
прорыв 

сезон: III

# КЕЙС

Центральный Банк Российской Федерации



Анализатор текстовых  
пресс-релизов: оценка  
взаимосвязи пресс-релиза  
и рейтинга



Министерство  
экономического развития  
Российской Федерации



# Кейсодержатель

Центральный Банк Российской Федерации

## 01 Сфера деятельности

Финансы

## 02 Краткое описание кейса

Оценка взаимосвязи между пресс-релизами кредитных рейтинговых агентств и присвоенными кредитными рейтингами по национальной рейтинговой шкале для Российской Федерации с учетом методологии.



### Сайт организации

[www.cbr.ru](http://www.cbr.ru)

# Постановка задачи

На основании исторических пресс-релизов кредитных рейтинговых агентств необходимо построить интерпретируемую ML-модель, устанавливающую взаимосвязь между текстом пресс-релиза и присвоенным кредитным рейтингом организации с учетом методологических особенностей оценки рейтинга. Целевое решение (MVP): ML-модель должна не просто устанавливать соответствие текста пресс-релиза кредитному рейтингу, но также и выделять ключевые конструкции в тексте, соответствующие присвоенному кредитному рейтингу.



Министерство  
экономического развития  
Российской Федерации



цифровой  
прорыв

сезон: III



# Проблематика

Законодательством Российской Федерации предусмотрено раскрытие кредитных рейтингов, сопровождающееся в том числе обоснованием присвоенного кредитного рейтинга или прогноза по кредитному рейтингу.

Так как Федеральный закон № 222-ФЗ направлен на обеспечение защиты прав и законных интересов рейтингуемых лиц и пользователей кредитных рейтингов, включая кредиторов и инвесторов, а также на обеспечение прозрачности и независимости деятельности кредитных рейтинговых агентств (КРА), пресс-релиз должен давать корректное и прозрачное обоснование уровней кредитоспособности рейтингуемых лиц.

# Проблематика

Сами по себе пресс-релизы представляют собой текст с указанием присвоенного рейтинга и прогноза по рейтингу, а также приведением ключевых факторов, объясняющих данное решение. Методология присвоения кредитного рейтинга пересматривается не реже 1 раза в год, при этом методологии каждого из 4х КРА, внесенных в реестр Банка России, различаются.

В рамках данной задачи допустимо пренебречь изменением методологии в рамках одного КРА, а также различием в методологических подходах разных КРА при определении рейтинга, при этом используя таблицу взаимно однозначного соответствия рейтингов между различными КРА, которую предоставляет кейсодержатель.

Наиболее сложной частью данной задачи является определение и визуализация ключевых словарных конструкций, сформулированных в пресс-релизе и напрямую влияющих на рейтинг.

# Решение

Решение кейса представляет собой интерпретируемая ML-модель, способная устанавливать зависимость между текстом пресс-релиза и присвоенным кредитным рейтингом с учетом методологии, результат работы ML-модели в виде построенных значений рейтинга для каждого пресс-релиза в датасете (и открытого, предназначенного для обучения, и закрытого, предназначенного для контроля), а также распознанные методологические факторы. Решение можно представить в виде веб-сервиса, способного определить уровень рейтинга в двух градациях (детализированной – 14 уровней и укрупненной – один из 7 уровней) для произвольного текста пресс-релиза, однако целевым будет применение данной модели в целях решения внутренних аналитических задач Банка России.

# Решение

Предпочтение будет отдаваться тем решениям, которые не просто дают точный результат, но также выделяют ключевые слова пресс-релиза, отражающие присвоенный кредитный рейтинг, и распознают факторы, указанные в пресс-релизе, как повлиявшие на итоговое значение присвоенного рейтинга.

В качестве итогового результата работы модели на тестовом датасете ожидается:

- уровень рейтинга для каждого текста из тестового датасета (в целевом и укрупненном виде);
- приведенные ключевые слова, повлиявшие на решение (наиболее релевантные);
- указание факторов и их значений, детектированных моделью в пресс-релизе (опционально);
- код обучения и результат inference ML-модели, использованной для генерации результатов по п.1, п.2 и п.3 на всем датасете (открытом и закрытом).

# Стек технологий, обязательных к использованию

Необходимые  
данные,  
дополнения,  
пояснения,  
уточнения



Министерство  
экономического развития  
Российской Федерации



цифровой  
прорыв

сезон: III

## 01

*Обязательно: Python*

*Предпочтительно: PyTorch/TensorFlow*

## 02

*Требования к железу: GPU 8GB*





# Оценка

→ Для оценки решений применяется метод экспертных оценок и автоматизированные средства оценивания.

→ Жюри состоит из отраслевых экспертов и/или представителей кейсодержателя.

→ На основании описанных ниже характеристик, жюри выставляет оценки 0-3 балла.

→ Итоговая оценка определяется как сумма баллов всех экспертов: технического, отраслевого и/или представителя кейсодержателя, как значение, выданное автоматизированными средствами оценивания, либо как итоговый балл жюри, умноженный на оценку автоматизированной системы.

# Технический эксперт оценивает решение по следующим критериям:

## 01

Запускаемость кода

## 02

Обоснованность выбранного метода (описание подходов к решению, их обоснование и релевантность задаче)

## 03

Точность работы алгоритма (возможность оценить формальной метрикой с обоснованием выбора)

## 04

Адаптивность/  
Масштабируемость

## 05

Отсутствие в решении импортного ПО и библиотек, кроме свободно распространяемого с обоснованием выбора

## 06

Наличие интеграционных интерфейсов, в первую очередь интерфейсов загрузки данных

Автоматизированные средства оценивания точности работы предложенных участниками алгоритмов (решений) выставляют оценку в диапазоне 0-1, где 1 равно 100% точности работы решения.

Итоговая оценка определяется как итоговый балл жюри, умноженный на оценку автоматизированной системы.

# Отраслевой эксперт и/или представитель кейсодержателя оценивает решение по следующим критериям:

## 01

Релевантность поставленной задаче  
(команда погрузилась в отрасль,  
проблематику; предложенное решение  
соответствует поставленной задаче;  
проблема и решение структурированы)

## 02

Уровень  
реализации  
(концепция/  
прототип и т.д.)

## 03

Проработка  
пользовательских  
историй (UX/ UI)

## 04

Реализация в решении  
требований Заказчика

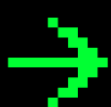
## 05

Выступление команды (умение  
презентовать результаты своей работы,  
строить логичный, понятный и интересный  
рассказ для презентации результатов своей  
работы)



# цифровой прорыв

сезон: III



Министерство  
экономического развития  
Российской Федерации

